

# Butterflies

Amal Almutairy

10/21/2020

## Butterflies:

```
butterflies <- read_csv("Butterflies.csv")
```

```
## Parsed with column specification:
## cols(
##   Site = col_character(),
##   Butterfly_N = col_character(),
##   wingspan = col_double(),
##   Speed = col_double(),
##   Weight = col_double(),
##   Origin = col_character(),
##   Color = col_character()
## )
```

## Descriptive statistics

Wingspan: Standard Deviation, Mean and Summary

```
## [1] 1.565508
```

```
## [1] 5.365
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      4.000   4.200   4.650   5.365   5.850   8.600
```

Speed

```
sd(butterflies$Speed, na.rm = TRUE) # is 4.186759
```

```
## [1] 4.186759
```

```
mean(butterflies$Speed, na.rm = TRUE) # is 15.65
```

```
## [1] 15.65
```

```
summary(butterflies$Speed)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      12.00   12.50   14.00   15.65   18.50   25.00
```

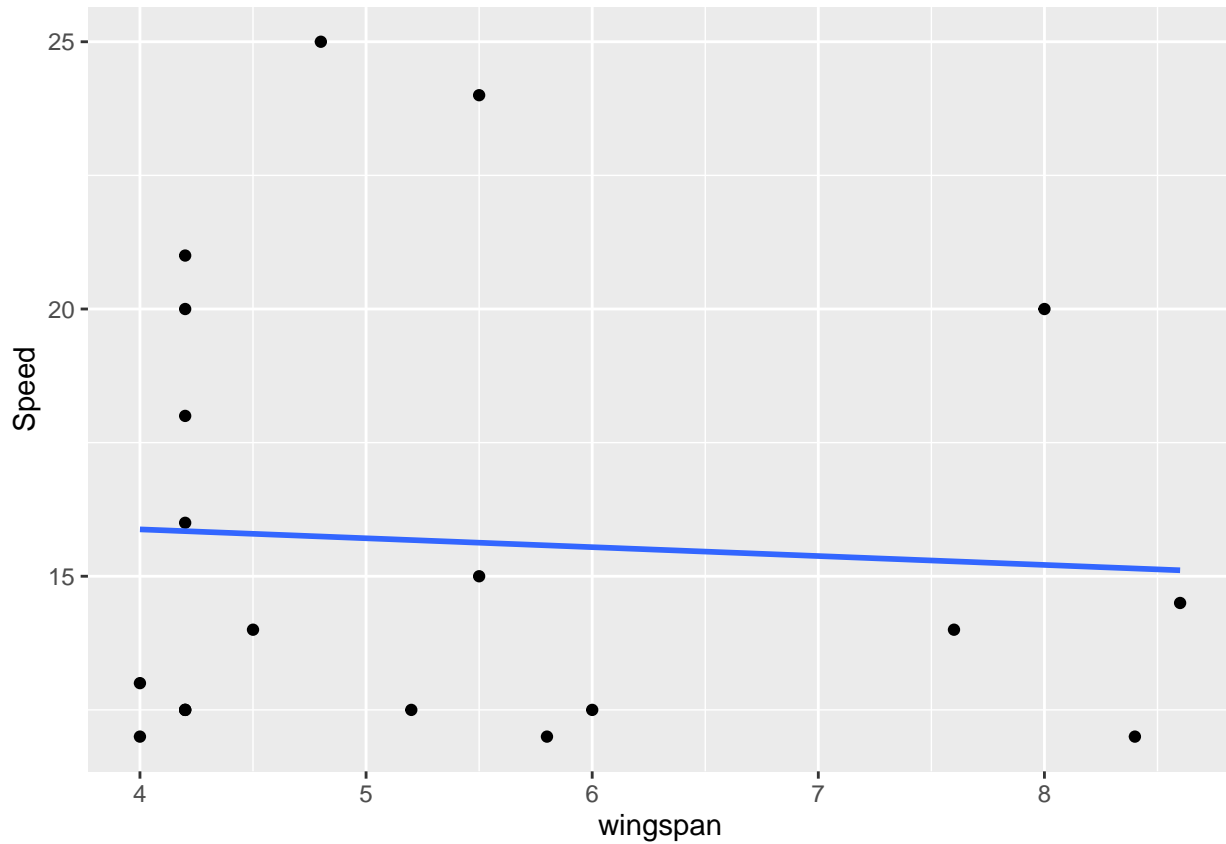
## Influencers of Speed

What are the greatest influencers of a butterfly's speed?

Does wingspan influent speed?

```
ggplot(butterflies, aes(x = wingspan, y = Speed)) +  
  geom_point() +  
  geom_smooth(se = FALSE, method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



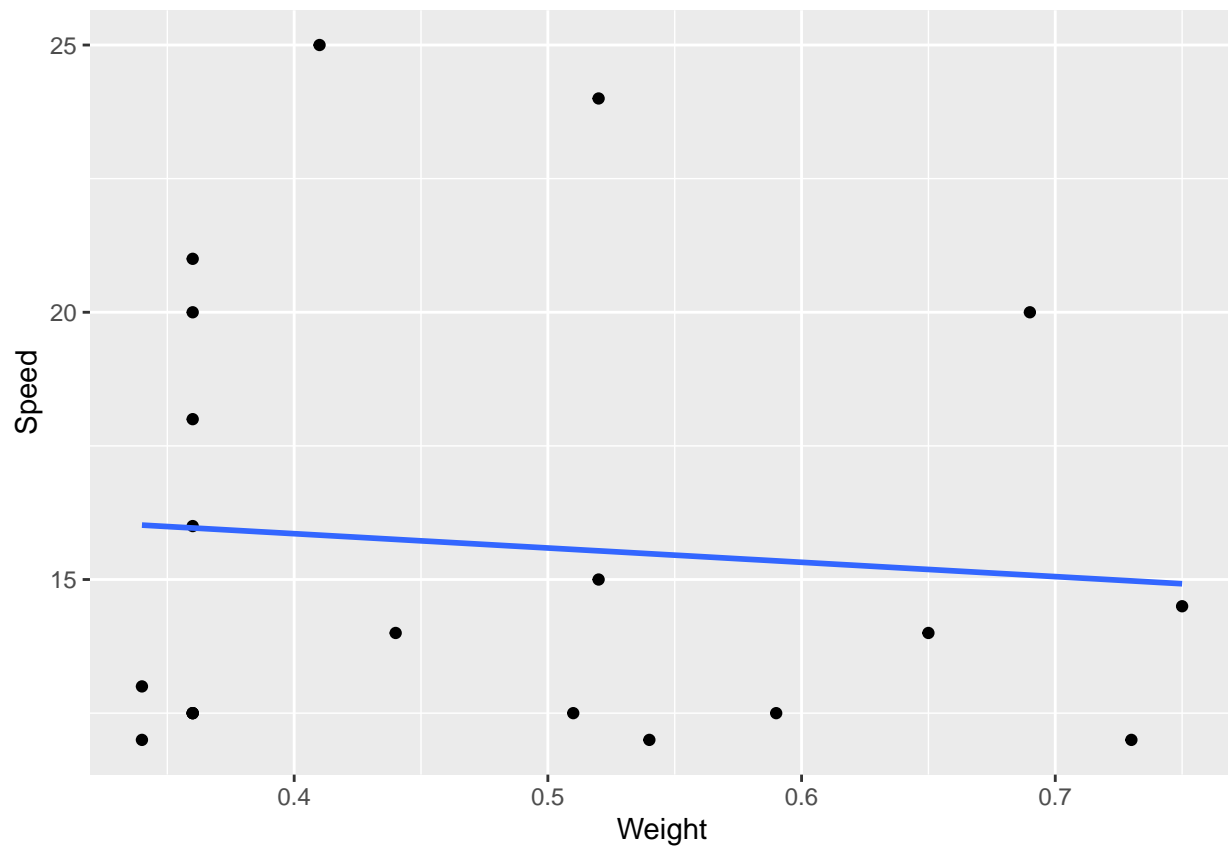
Describe this quantitatively? i.e.

- What is the model?
- Is it significant?
- Is the model appropriate for the data?
- Can you define a different model?

The other continuous variable would be weight?

```
ggplot(butterflies, aes(x = Weight, y = Speed)) +  
  geom_point() +  
  geom_smooth(se = FALSE, method = "lm")
```

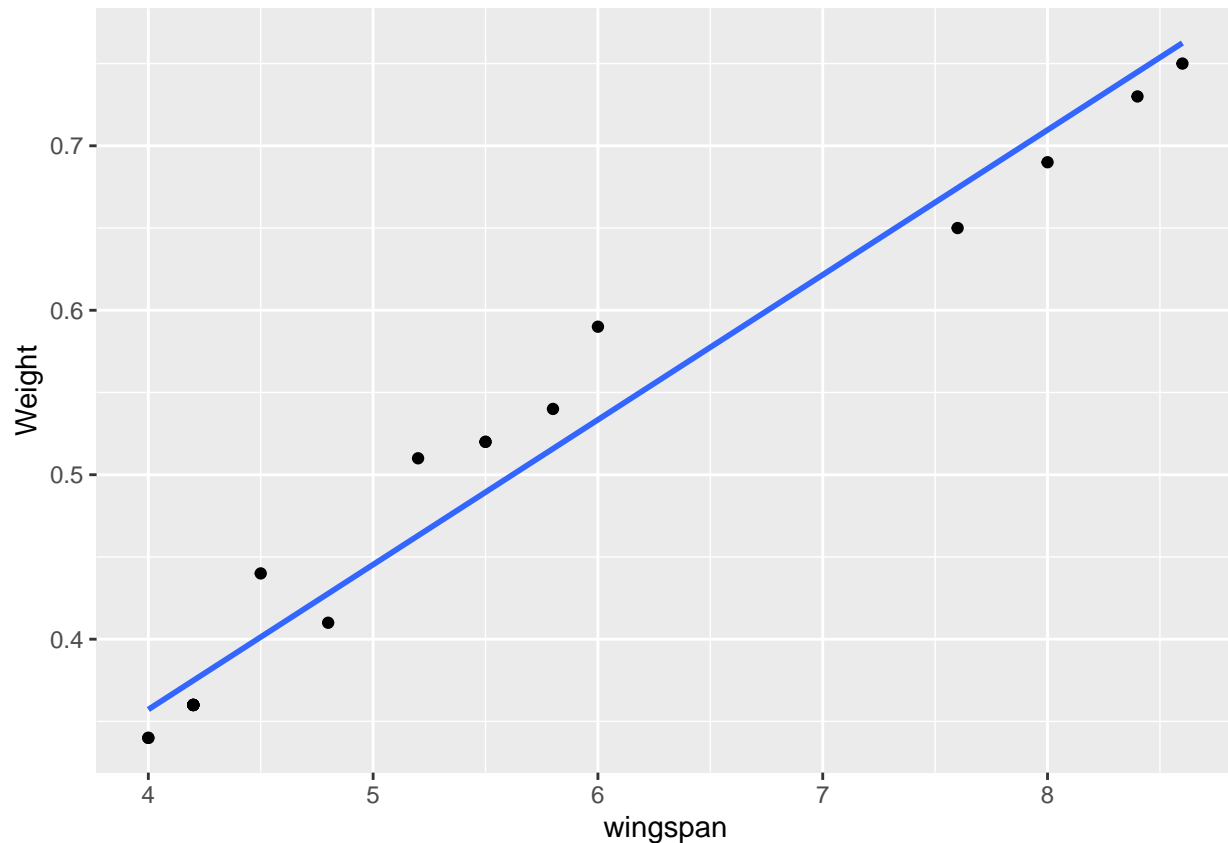
```
## `geom_smooth()` using formula 'y ~ x'
```



Is there a combinatorial effect of weight and wingspan on the speed of a butterfly?

```
ggplot(butterflies, aes(x = wingspan, y = Weight)) +  
  geom_point() +  
  geom_smooth(se = FALSE, method = "lm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```



So weight and wingspan are “co-linear” so it’s not surprising that weight doesn’t predict speed, since wingspan, we already doesn’t.

## Inferential Statistics

```
# Inferential Statistics
Butterflies_lm <- lm(wingspan ~ Butterfly_N , data = butterflies)

Butterflies_lm %>%
  anova(Butterflies_lm)
```

```
## Analysis of Variance Table
##
## Model 1: wingspan ~ Butterfly_N
## Model 2: wingspan ~ Butterfly_N
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1      16 1.6713
## 2      16 1.6713  0          0
```